Winner Determination of Open Innovation Contests in Online Markets

Completed Research Paper

Yang Yang **Pei-yu Chen** **Rajiv Banker**
Fox School of Business
Temple University
Philadelphia, PA U.S.A
yangyang@temple.edu pychen@temple.edu banker@temple.edu

Abstract

Online innovation contests have been used by more and more firms for idea seeking and problem solving. Most studies of contests take the perspective of innovation seekers, and little is known about solvers' strategies and responses. However, contest performance also relies on understanding solver responses. This paper provides insights to these questions. Specifically, we show that past experience of a solver is a good predictor of his future winning probability and that winners are more likely to be those who submit early or later during the submission period as opposed to those submit in the middle. We also find that "strategic waiting" (to submit solutions) is associated with higher winning probability. Furthermore, we show that different contests appear to attract solvers with different expertise, which invalids the common assumption of fixed solver expertise distribution across projects in previous literature. This finding has strategic implications to the design of contest parameters.

Keywords: Open Innovation, Winner determination, Strategic bidding, Online contest, Online bidding, Crowdsourcing, Past Experience

Introduction

Open innovation contest (Terwiesch and Xu 2008; Yang et al. 2009), or *crowdsourcing contest* (Archak and Sundararajan 2009; Howe 2006; Yang et al. 2008), in which an innovation seeker, which can be a firm, an organization or an individual, runs a contest to seek innovative ideas or solutions to a problem, has been adopted by more and more firms for problem solving and new product development. By launching online innovation contests, firms can easily reach large volume of external solvers with diversified background. A larger pool of potential solvers can help facilitate faster and potentially better ideas or solutions compared to internal innovation efforts. For instance, in September 2008, Google funded a $10M launch of an open innovation contest. The project was called Project 10^100 and calling for ideas to change the world, in the hope of making world better. Since 2006, Netflix set 1M prize every year seeking to substantially improve the accuracy of predictions about how much someone is going to enjoy a movie based on their movie preferences. In addition to self-facilitated open innovation contests, several markets also exist to facilitate open innovation contests. InnoCentive, founded in 2001, is the first online market to host open innovation projects in form of contests (Allio 2004). It was originally built to facilitate seeking for innovative medicine solutions. But for now, as an emerging result, a variety of projects are posted there, ranging from website LOGO design, algorithm design to complex project such as construction design. There also exist other contest markets with different focus or geographic region, such as Topcoder, and TaskCN, and various organizations or companies are using these markets as platforms for open innovation projects. For firms lacking IT deployment capacity or channels of reaching huge potential solvers, launching their innovation contests in a mature online market is an obviously better choice. In the future, we expect that more and more firms will adopt open innovation to mitigate the risks of internal projects and identify exceptional solutions from across the globe.

To launch an online contest, an innovation seeker needs to post project details including a set of fixed prize, duration, description and etc. All potential solvers decide whether to join the contest or not and if they join, they compete by submitting their solutions/products. After the contest ends, the seeker will assign the prize to the solver who provides the best idea/solution according to some criteria. Several theoretical studies have been done on contests. Dahan and Mendelson (2001) pursuit an extreme value model and argue that the final performance of a product development contest is decided by the top tier distribution of solvers. Terwiesch and Xu (2008) have extended this model to more general contest situation where projects may have multiple dimensions such as expertise and ideation based. If the distribution of solver ability is irrelevant to the project characteristics, having more solvers will give a seeker higher chance to get better solutions. A typical assumption made in previous theory literature is that all solvers receive the same information and compete simultaneously.

However, in most online contest markets (i.e. Zhubajie, TaskCN, 99designs, CrowdDesign and etc.)[1], we observe that solvers are competing dynamically instead of simultaneously. As shown in the typical timeline of an online contest (Figure 1), a solver can enter the contest at any time from the time the contest starts to when it ends. Since an online contest market usually makes the status of a contest (including the number of submissions, and even their final submissions, etc.) available, therefore, depending on the time the solver enters, they receive different information and can take different strategies accordingly. For example, by waiting longer, a potential solver has more information regarding number of competing solvers and the quality of submissions available. As a result, he[2] has better evaluation of his chance of winning and can take more effective action accordingly. On the other hand, by submitting a good solution early, the solver may discourage other solvers from submitting solutions further. Besides, submissions are usually accessible to public and this policy makes each solver consider when to submit his solutions to maximize his winning probability. Thus, the winning result is also impacted by solvers' bidding behaviors.

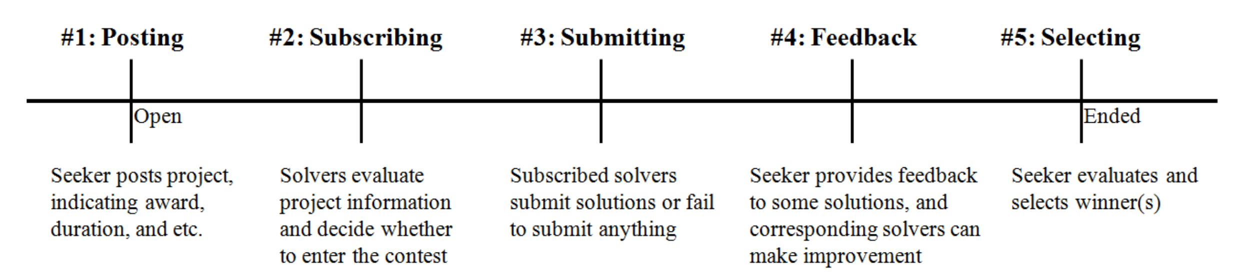

Figure 1. Dynamic Timeline of Online Contest

The goal of having an online contest is to attract high quality solvers and obtain good and diversified solutions (Terwiesch and Ulrich 2009; Terwiesch and Xu 2008), which relies not only on contest design parameters but also on the understanding of solvers' incentives and strategic actions. A good contest

[1] We consider the contest scenario of most popular online contest markets. Although InnoCentive receives lots of academic attention, it is not a popular online market. Thus its business model and policies are not considered in our study.

[2] In economic research, for the convenience of discussion, buyer is usually "she" and seller is "he". Similarly in this paper, we call a seeker "she", and a solver "he".

design should take into account solver behavior, and it is important to understand the strategic interactions between solvers. However, to our knowledge, not only there is scant empirical literature on contests, but there is a lack of understanding of how solvers compete with one another strategically. This research has two goals: First, we aim to unveil the factors that can influence a solver's chance of winning. Second, perhaps as a result of lack of understanding of how solvers compete, a common assumption made in previous theory literature is that solvers are drawn from the same distribution regardless of contest design parameters. That is, solver distribution is the same across contests (Terwiesch and Xu 2008). In reality though, there are often more than one contest, especially on a contest market, that a potential solver can participate in, and due to capacity constraints, solvers may self-select themselves into different contests. That is, it is likely that some contests are more able to attract more experienced solvers than others. If this is true, then previous assumption underlying many theory works is invalid. We are interested in testing whether this is empirically true or not.

The remainder of this paper is organized as follows. In next section we discuss the features of online contest markets that may affect solvers' performances, and hypotheses are developed. Then we describe our data source, variable measurement, and the empirical models to test our hypotheses. At last we present our empirical results, followed by conclusions and implications.

Theory and Hypotheses

This study takes the perspective of contest solvers. Consider an innovation seeker posting a project on an online contest market and a number of solvers enter. To a solver, how to maximize his probability of winning is perhaps the question he is most concerned about. The quality v_i of solution from solver i, where i =1... n can be modeled in a linear format (Terwiesch and Xu 2008):

$$v_i(\beta_i, e_i, \xi_i) = \beta_i + r(e_i) + \xi_i \tag{1}$$

where β_i is the expertise level of solver i. It is usually assumed that the distribution of expertise is known and it is fixed across all contests. $r(e_i)$ is the output of effort when solver i executes effort e_i. $r(e_i)$ is increasing with e_i. ξ_i is a random error term of each solution. This random error also captures the unobserved preferences of the seekers and includes the diversified ideation-based output. Since β_i is fixed, the variance of a solver's performance is mainly based on the effort output r(e_i) and random error. As noted earlier, due to dynamic competition process, solvers' bidding strategies may also impact the winning results, thus equation 1 can be modified to:

$$v_i(\beta_i, e_i, s_i, \xi_i) = \beta_i + r(e_i) + s_i + \xi_i \tag{2}$$

Next we will explore the impacts of a solver's expertise and bidding strategies on winning results. Effort output r(e_i) is usually difficult to observe directly but its impact will be discussed with related bidding strategies.

Expertise/Past Experience Impact

Previous studies indicate that individual performance is restricted by his expertise (Banker and Iny 2008; Snir and Hitt 2003; Terwiesch and Xu 2008). How to define expertise with online contest scenario in a quantifiable way is a key task for empirical study. In organizational literature, there are two views of expertise. The first view is based on the knowledge possession of individual (Rorty 1979). In this approach, knowledge is a material that can be abstracted, explicitly represented, codified, and accessed (Walsh 1995). In online contest, free entry attracts solvers with diversified background. Following this view, Terwiesch and Xu (2008) suggest that expertise is usually a measure of a solver's past experience and knowledge for a particular problem. For example, the winner of a LOGO design contest is more likely to be a designer than a chemist. However, diversified expertise is hard to reach a unified define for crowd of solvers, because one area of knowledge is not comparable to another. Another perspective perceives expertise is context dependent, emerging from patterned interactions and practices in specific scenarios (Brown and Duguid 1991; Faraj and Sproull 2000; Snir and Hitt 2003). Following this view, expertise of a solver can be reflected by his emergent and relative contest performance results by comparing with other solvers in the same community such as participation experience and past experience.

In reverse auction literature, past experiences have been proved to be good predictors for future winner determination (Banker and Iny 2008). In reverse auction, agents are bidding with project proposals which are signals of final solutions. An experienced agent gaining high value of winning propensity and feedback represents that this agent is consistently good at negotiating and maximizing principal's utility by sending "right signals". In online contests, solvers are bidding with final solutions. In other words, solvers are bidding with their skill of doing, not negotiating. A seeker with better performance discloses his consistent better expertise. On one hand, we could conclude that this solver has higher expertise level, and on the other hand, we can predict that this solver will have higher probability of winning. Terwiesch and Ulrich (2009) have done an experiment with MBA students and provide proof to show that the quality of submissions of a specific solver is consistent over time. Therefore, we expect that expertise level is positively associated with winning probability in the future. We have the following hypothesis:

H1: A solver with higher expertise level is more likely to be the winner.

Temporal Strategy

Internet enabled initiatives allow geographically distributed players to collaborate or compete together. Comparing to traditional or offline contest, one distinguished feature is that online contests allow players to compete dynamically, instead of simultaneously. When to enter and bid is a common concern of all solvers. A long stream of studies has been done to understand auction bidder's dynamic bidding strategy and benefits. Late bidding in Internet auctions has attracted a good deal of attention (Ockenfels and Roth 2001; Ockenfels and Roth 2006; Roth and Ockenfels 2002; Vadovic 2009). The intuition behind last-minute bidding in a private value auction is that there is an incentive not to bid high when there is still time for other bidders to react, to avoid an early coming bidding war that will raise the expected final transaction price (Ockenfels and Roth 2001). On the opposite side, an early and high bid can lower bidder's searching cost for substitutions and make other competitors less interested in this competition (Vadovic 2009). Unfortunately little attention has been paid to strategic bidding of online contest. In an online contest, with project search tool, a solver may strategically choose when to enter and when to submit his solution, as long as the contest is not ended. From a solver's perspective, his most interest is to know the best strategy that maximizes his probability of winning. To know that, we need understand all the benefits and drawbacks of early and late actions, in order to find out the dominant strategy or any equilibrium in this dynamic game.

Benefits mean any positive impacts on the goal of winning, and drawbacks are for the opposite impacts. Generally speaking, by entering early, the solvers can have option of when to submit and which, at least, is better than no options. So entering earlier is always not a bad choice. Although entering early requires the solver be more patient in waiting for the result, this strategy has no negative impact on winning result. By submitting the solution early, the solver has the advantage of getting feedback from the seeker earlier and still having time to implement the feedback accordingly. Successfully implementing the feedback seeker desires will undoubtedly increase the solver's probability of winning. In addition, by submitting a good solution earlier, the solver may also discourage other solvers from competing further because a good solution may make other solvers perceive lower winning probability and therefore drop out. On the other hand, by purposely waiting, the solver has more information about the number of competing solvers and the quality of submissions. More information gives the solver better evaluation of winning probability, and the solver can then take actions accordingly. For example, they will pursue the project further if he perceives high probability of winning and drop out if winning probability is low. This suggests that a solver will incur costs only when he perceives high probability of winning, reducing the expected costs However, the fact that a solver submits late may be a direct result of entering the contest late. Therefore, the time lapse between entering time and submission time may reveal something strategic. Specifically, a solver who purposely waits will have relatively higher time lapse compared to solvers who submit late because of entering the contest late.

Let's take a simplified participation process as example. Assume that a solver can only choose to enter or submit at two moments, called *early*, and *late*. After entered, he needs further decide when to submit. If he enters early, he can choose to submit early or late. If he enters late, he can only submit late. In total there are three strategies of participations. We summarize the benefits of each strategy in table 1.

Table 1. Temporal Strategy Analysis of Solvers

		Submit Early		Submit Late
Enter Early	I	✓ More likely to receive feedback and make improvement ✓ Stop others to enter by submitting a high quality solution ✓ More time to work on solution ✓ More chances to learn from competing solutions.	II	✓ Less likely to be copied by others ✓ More time to work on solution ✓ More chances to learn from competing solutions.
		× More likely to have similar competing solutions		× Less likely to receive feedback
Enter Late			III	✓ Less likely to be copied by others
				× Less likely to receive feedback × Lost benefits of entering early

Notes: ✓- benefit × - drawback

By comparing strategy II to strategy III, it is obvious that strategy II has more benefits than strategy III as it gives the solver more time to work on solutions and more chances to learn from competing solutions. In other words, strategy II dominates strategy III.

Comparing strategy I to strategy II, after removing all the benefits in common, we can see that each of them still have some unique benefits. Strategy I solver is more likely to receive feedback and make improvement desired by the seeker; he may also prevent others from entering by submitting a high quality solution. Strategy II solver is more likely to submit a unique solution and make the solution more competitive. In perception of solvers, it's not clear which strategy is dominant. Since there are no dominant strategies in the dichotomous contest, it is not surprising that there exist more than a unique equilibrium.

Above analysis is based on a simplified situation where only two entry/submit moments are available. In a real contest, solvers can submit at any moment when the contest is open. At any moment, each solver has the benefit of submitting earlier than some solvers and the benefit of submitting later than the other competing solvers. We assume the benefit of submitting early $P_{submit\ early}$ at moment t is convex and monotonically decreasing (i. e., $P_{submit\ early}|_t = C_1 + C_2 \exp(-C_3 t)$, where C_1, C_2, and C_3 are constants and C_2, C_3 >0). When a contest is just launched, $P_{submit\ early}$ is the highest. When a contest is ending, $P_{submit\ early}$ is the lowest. Similarly, we also assume the benefit of submitting late $P_{submit\ late}$ at moment t is convex and monotonically increasing (i. e., $P_{submit\ late}|_t = C_4 + C_5 \exp(C_6 t)$, where C_4, C_5, and C_6 are constants and C_5, C_6>0). When a contest is just launched, $P_{submit\ late}$ is the lowest. When a contest is ending, $P_{submit\ late}$ is the highest. The overall benefit of submitting at moment t is $P|_t = P_{submit\ early} + P_{submit\ late}$, which is convex and non-monotonic. In other words, either submitting very early or late may increase the chance of winning than those submissions in between. Thus we have:

H2a: the probability of winning follows a U shape in terms of solvers' relative submission order.

In a dichotomous contest, entering early is always better than late, and this result is still hold in a real contest. In a real contest, after a submission is observed, we also concern about the interaction between entering time and submitting time. To capture entering strategy and the interaction with submitting strategy, the time lapse between the entering and submission time is more interesting. This time lapse reveals not only how early a solver entered, but also the strategic waiting before final submission. On another hand, it may also be a noisy measure of effort output r(e_i) in equation 2. Regardless of strategic waiting or higher efforts, we expect that a solver's chance of winning is increasing in this time lapse. We have:

H2b: The longer time lapse between entering time and submission time is associated with higher chance of winning.

Strategic Choice of Projects

As noted earlier, another question of our interest is to test whether solvers with different expertise level choose different projects systematically. For example, do solvers with higher expertise tend to choose higher-prize projects? Many previous studies assume that expertise has no impact on choice in order to simplify the theoretical model. However Terweisch and Xu (2008) doubt the validity of this common assumption. In online reverse auction market. Snir and Hitt (2003) find that high value projects attract pools of solvers with lower quality in average. Does contest market have similar phenomenon? An experienced solver should have learned how to increase his profit by choosing "right" projects. If so, this would suggest some self-selection bias and invalid the common assumption of previous studies that expertise distribution are constant across different projects. To test the validity of above common assumption, we have the following null hypothesis:

H3: A solver's expertise does not affect his choice of projects.

Data and Estimation Model

In this part, we introduce our research site, sample selection, sample data, measurement of constructs, and estimation model.

Research Site and Sample Selection

We collect data from TaskCN.com, which is one of the largest online service markets in China and founded in 2005. By the end of 2009, TaskCN has over 2.7 million registered solvers. This market allows anyone to launch a contest with prize deposit in advance. All contests are free to enter. This contest model has also been adopted by most online contest markets such as Zhubajie, TopCoder, DesignCrowd and etc.

We collected all contests launched between June 2006 and October 2008. In total, there are 7,728 contest projects. Around 20% of the projects are multi-winner projects. We eliminated these projects since the optimal design of prize structure has not been fully understood and is not the core of this study. Sales force contests were also eliminated since sales force contests are pursing maximizing the overall performance of all solutions, not just one or several best solution (Liu et al. 2007; Terwiesch and Xu 2008). After censored, our sample includes 1995 contests with 216,812 submissions.

Sample Data and Measurement

Expertise Variables

Following perspective that expertise is context dependent, emerging from patterned interactions and practices in specific scenarios (Brown and Duguid 1991; Faraj and Sproull 2000), in online contest market, expertise of a solver can be reflected by his emergent and relative contest performance results by comparing with other solvers in the same community such as participation experience and past experience. In practice it's hard to define expertise objectively. In our study, a solver's participation experience and past experience can be measured by[3]: his number of participations *N_Join*, number of winnings *N_Win*, membership age *MembershipAge*, and winning propensity *WinPerJoin* prior to entering a new contest. Since *WinPerJoin = N_Win / N_Join*, we only include two of three at the same time. WinPerJoin is an ability measure which is unique and critical. Although participation experience is important, the winning experiences can tell a solver how to win, so we keep N_Win. Some solvers have very long membership age, but these extremely high levels of experiences are not likely to have proportionately additional credit for probability of winning. Hence, we take membership age natural-log transformed to reflect the diminishing marginal benefit of increased experience. In summary, in our estimation model, we use *N_Win*, Ln*(MembershipAge)*, and *WinPerJoin* to proxy for expertise measures.

Table 2A. Expertise Variables Definition

Variable	Definition
N_Win	Total number of winnings before current contest.
N_Join	Total number of participations before current contest
WinPerJoin	N_Win / N_Join, winning propensity
MembershipAge	Number of days between registration and entering current contest

Table 2B. Descriptive Statistics of Expertise Variables

Variable	Mean	Std Dev.	Min	Max
N_Join	19.992	66.933	1	1368
N_Win	1.032	5.916	0	163
MembershipAge (Days)	85.4778	142.31	0.0001	813.16
WinPerJoin	0.018	0.070	0	1
No. of Observations	216,812			

Temporal Strategy Variables

The strategies that a solver can fully control are limited. We list all temporal variables in Table 3A. As discussed before, the most concern of a solver is when to join *(JoinTime*) the contest and submit to his solution (*SubmitTime)*. In order to compare across contests with different duration settings, we standardize these time values with following equation:

$$G\_SubmitTime = \frac{SubmitTime - StartTime}{Duration} \quad (3)$$

G_SubmitTime is a value between 0 and 1. If *G_SubmitTime=1*, it means the solver submitted just before the project is closed. With the same method, we get *G_JoinTime* and *G_ ⊿Time*. *G_JoinTime* measures the standardized time that a solver enters a contest. *G_ ⊿Time* measures the standardized time lapse

[3] In auction studies, feedback rating is a key indicator of expertise (Snir and Hitt 2003; Banker and Iny 2008). However in contest, feedback rating doesn't help since nearly all winners are receiving positive feedbacks. That's because winners are chosen based on final performances, and there is no information problem. A winner is always relatively better than other competitors. Our data shows that 98% feedbacks are positive, and nearly no variances.

difference between *JoinTime* and *SubmitTime*. *G_ ΔTime* captures not only when to submit, but also the interaction with submission time. So this value is more informative than *G_JoinTime*.

However the standardized time value may not reveal a solver's temporal strategy. In practice, if there is only one solver in pool, no matter when he submits, he is always the first and the last to submit. In this circumstance, *JoinTime* and *SubmitTime* are useless in winner determination. To reflection the variation of competing strategy of relatively "early" or "late" submissions, the submission order of a solver (*SubmitOrder*) is more important. We define *G_SubmitOrder as*:

$$G\_SubmitOrder = \frac{SubmitOrder}{N\_Submit} \tag{4}$$

G_SubmitOrder is also a ratio between 0 and 1. If *G_SubmitOrder=1,* it means this solver was the last one to submit. With the same method, we have *G_JoinOrder*, which measures the relative time that a solver enters a contest.

Table 3A. Strategy Variables Definition

Variable	Definition
JoinTime	The time that a solver declares that he will join the competition
G_JoinTime	(JoinTime-StartTime)/Project Duration
JoinOrder	The number of joined solvers when this solver joins.
G_JoinOrder	JoinOrder/N_Submit
SubmitTime	The time that a solver submits his initial solution.
G_submitTime	(SubmitTime-StartTime)/Project Duration
SubmitOrder	The number of submitted solutions when this solver submits.
G_SubmitOrder	SubmitOrder/N_submit
ΔTime	Submit Time – Join Time
G_ΔTime	ΔTime/Duration

To test H2a and H2b, *G_SubmitOrder* and *G_ΔTime* are needed. Descriptive statistics of these two are listed in Table 3B.

Table 3B. Descriptive Statistics of Strategy Variables

Variable	Mean	Std Dev.	Min	Max
G_SubmitOrder	0.500	0.284	0	1
G_ΔTime	0.051	0.131	0	0.9995
Number of Observations	216,812			

Correlation coefficients among all necessary variables that estimation model will use are listed in Table 3C.

Table 3C. Correlation Coefficients of Participation Variables

	N_Win	WinPerJoin	Ln(MembershipAge)	G_ΔTime	G_SubmitOrder
N_Win	1.000				
WinPerJoin	0.237	1.000			
Ln(MembershipAge)	0.216	0.165	1.000		
G_ΔTime	0.051	0.056	0.095	1.000	

G_SubmitOrder	- 0.035	- 0.022	- 0.016	0.282	1.000

Project Group Variables

Participations are grouped by contest projects. There are several project group variables which are fixed to each submission within a project. These variables may impact a solver's perceived probability of winning such as prize, duration, category, dimensions and etc. Descriptive statistics of these fixed project variables and categorical means are in Table 4 and 5.

Table 4. Descriptive Statistics of Fixed Project Variables

Variable	Mean	Std Dev.	Max	Min
Number of Solver	111.67	319.94	4498	1
Prize Amount (￥)	287.15	361.952	5500	1
Contest Duration (days)	21.28	15.81	104.11	0.0021
Number of Projects	1995			

Table 5. Category Means and Features of Projects

Category	Prize(￥)	Duration	No. Solvers	Percentage	Dimensions
Web Building	285.01	20.81	13.53	3.3%	Ideation & Expertise
Translation	67.58	13.37	118.78	1.40%	Expertise
Software Develop	126.84	17.32	8.31	5.45%	Expertise
Q&A	61.70	16.67	22.48	5.15%	Q&A
Media	316.89	16.05	15.49	3.25%	Ideation & Expertise
Web Design	483.57	21.66	18.19	6.65%	Ideation & Expertise
Product Design	545.55	25.32	18.87	5.15%	Ideation & Expertise
Packaging Design	439.73	30.43	31.23	1.9%	Ideation & Expertise
LOGO Design	358.66	22.59	60.56	40.5%	Ideation & Expertise
Interior Design	245.17	17.24	21.46	0.85%	Ideation & Expertise
Graphic Design	288.35	19.56	38.3	13.5%	Ideation & Expertise
Creative Writing	159.21	19.74	71.34	7.5%	Ideation
Naming	136.28	25.32	729.22	9.6%	Ideation

Estimation Model

To estimate what kind of solvers tend to win and to test H1 and H2, we employ the conditional logit model (Wooldridge 2001, Greene 2002) which is commonly employed in consumer choice research, especially when the number of alternative choices is large. For an innovation seeker of contest project *i*, her utility function of choosing solver *j*'s submission is:

$$U_{ij}^{*} = X_{j}\boldsymbol{\beta} + \xi_{ij} \quad j = 1,2,\ldots N_{i} \tag{5}$$

Where ξ_{ij} accounts for all unobservable attributes that can affect seeker's preferences. The value provided by solver *j* is determined by X_j, which include solver *j*'s expertise and temporal strategy with a common parameter vector $\boldsymbol{\beta}$. Although we cannot observe and measure the utility of each submission, we can conclude that the winning submission offers the highest utility to the seeker. That is, if the seeker finally chooses submission of solver w_i ($w_i \in \{1,2,...N_i\}$), then we can assume that this utility $U_{ij} \mid j = w_i$ is the maximum among N_i utilities. When N_i disturbances, ξ_{ij} is independently distributed with the Type I extreme value distribution (McFadden 1974). Then the probability that w_i is the only winning submission of project *i* is:

$$prob(W_i = w_i) = \frac{\exp X_j \boldsymbol{\beta}}{\sum_{j=1}^{N_i} \exp X_j \boldsymbol{\beta}}, \quad w_i \in \{1,2,...N_i\} \tag{6}$$

To estimate the coefficient vector $\boldsymbol{\beta}$, we use maximum likelihood criterion. Variable Win_{ik} is 1 if solver *k* is the winner, otherwise, Win_{ik} =0 for all non-winners. We estimate the $\boldsymbol{\beta}$ by maximizing the log-likelihood function:

$$\ln L = \ln \prod_{i=1}^{1995} (W_i = w_i) = \sum_{i=1}^{1995} \sum_{k=1}^{N_i} Win_{ik} \ln \Pr ob(W_i = w_k) \tag{7}$$

Results and Analysis

Table 6 reports the results of who is likely to be winner of a contest based on conditional logit model. Pearson correlation coefficients between each term are all lower than 0.3 (Table 3).

Table 6. Conditional Logit Model and Results

$$U_{ij}^* = \beta_{1a}(N\_Wins) + \beta_{1b}(MembershipAge) + \beta_{1c}(WinPerJoin) + \beta_{1d}(G\_SubmitOrder) + \beta_{1e}(G\_SubmitOrder)^2 + \beta_{1f}(G\_\Delta Time) + \xi$$

Variable	Symbol	Coefficient	Significance	Odds Ratio
N_Win	β1a	1.0442***	<0.001	2.8413
MembershipAge	β1b	- 0.9642	0.219	0.9080
WinPerJoin	β1c	0.1794	0.222	1.1965
G_SubmitOrder	β1d	- 2.0205***	<0.001	0.1325
G_SubmitOrder2	β1e	2.7364***	<0.001	15.4313
G_ΔTime	β1f	1.9991***	<0.001	7.3824
Number of observations	21,6812 observations grouped in 1995 cases			
Pseudo R2	8.84%			
LR Chi2	1671.91*** (Prob>Chi2 is less than 0.0001)			

Note: *~p<0.1, *~p<0.05, ***~p<0.01. Expertise variables have been transformed to percentile scores

Since we are using conditional logit model, the influence of each factor needs to be explained with significance and odds ratio. Above result shows that in our selected expertise variables, the winning experience is a good predictor of future winning result. Odds ratio of N_Win is larger than 1, so a solver with more winning experiences is more likely to win in the future. H1 is supported.

Regarding the temporal strategy, coefficients of *G_SubmitOrder* and *G_SubmitOrder*^{2} suggest a U shape curve (Figure 2) of winning probability. In other words, very early submissions and very late submissions have higher probability of winning than those in-between. Therefore, H2a is supported.

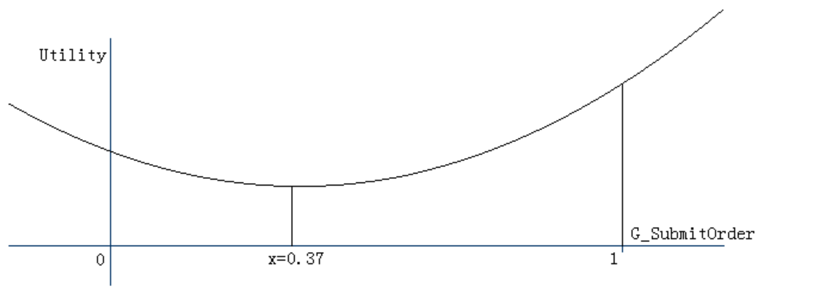

Figure 2. Estimated Utility – U shape

Besides, larger *G_ΔTime* is highly associated with higher probability of winning. So H2b is also supported.

We should note that one should be cautious in interpreting the results on temporal strategy. Specifically, it doesn't mean submitting earlier or later can definitely increase probability of winning. The intrinsic incentive of doing that should be receiving feedback earlier, discouraging other solvers from further pursuing the projects, or learning more competing submissions to make own submission more competitive. If a solver just intends to submit first or the last without concerning seeker's feedback or learning from competing submissions, his probability of winning won't increase at all. Our finding does suggest that more capable solvers may have self-selected themselves into submitting earlier or later. For example, those who submit late must be good ones and have perceived high winning probability because they have observed all available submissions and they would not pursue further unless he believes that he can beat others. On the other hand, a capable solver also has a lot to gain by submitting early because he can discourage other solvers from participating further while taking advantage of likely feedback to further improve his solution. This has some interesting implication to seekers. For example, when evaluating submissions is costly, a seeker may want to focus more on earlier and later submissions than in-between ones.

Given that the project characteristics are fixed within each contest, it is meaningless to add these variables in conditional logit regression. In order to know how these impacts vary for different projects, we get sectional results of project categories, dimensions, prize values, and duration sections. Winning ratio and membership age has been removed due to insignificance. Income is also removed from regression since it's out of interest here.

Table 7. Sectional Results by Category

Variables	LOGO Design (647 Cases)	Graphic Design (235 Cases)	Naming (163 Cases)	Software (82 Cases)
N_Win	0.0332***	0.0168*	0.0075	- 0.0182
G_SubmitOrder	0.7780	- 1.3345	- 5.6585***	1.7127
G_SubmitOrder^2	0.1653	2.2804**	5.5533***	- 0.9925
G_ΔTime	2.5350***	2.2611***	1.8647***	4.2145***

Pseudo R^2	13.33%	9.73%	3.32%	16.26%

Note: *~p<0.1, *~p<0.05, ***~p<0.001.

Table 8. Sectional Results by Dimension

Variables	Ideation	Expertise	Ideation & Expertise	Q&A
	(342 Cases)	(137 Cases)	(1413 Cases)	(93 Cases)
N_Win	0.0030	- 0.0186	0.0207***	- 0.0046
G_SubmitOrder	- 4.4475***	- 1.0155	0.0137	- 3.9954**
G_SubmitOrder²	4.6086***	0.9489	1.0332**	2.5395*
G_ΔTime	1.8872***	4.0297***	2.5415***	2.9361**
Pseudo R^2	3.38%	11.87%	11.62%	8.07%

Note: *~p<0.1, *~p<0.05, ***~p<0.01.

Result in Table 7 shows that submitting strategies show no significant impacts on winning results of LOGO design, software development projects and all expertise based projects, but very significant impacts to naming or pure ideation based projects. For ideation based projects, the probability based on submitting order is a symmetric U shape. In other words, the first submission and last submission have equal chance of winning and higher than other submissions. However in Table 8, *G_SubmitOrder* is not significant and *G_SubmitOrder²* shows positive and significant impact. So for ideation and expertise based projects, winners are more likely from those who submit late. Longer *G_ΔTime* is always helpful to all kinds of projects. Expertise variables also work differently for different categories. For software projects or all expertise based projects, previous performance doesn't appear to help. This is surprising since the winning result is assumed to be consistent with expertise level. Both expertise and strategic variables behavior quite different influence from category to category, so project category as a project characteristic can moderate expertise and strategic impact on probability of winning.

To understand how variances of contest prize and duration moderate those impacts, we split all projects into several sections: prize higher or lower than average, duration longer or shorter than average (Table 9). Results show that when prize is higher, expertise variables play more important role in predicting winners, and winners are also more likely to be those who submit late. When duration is shorter, expertise plays less important role. But again, those who submit later are more likely to be winners. In other words, for high prize and short duration projects, learning from other submissions are more beneficial.

Table 9. Sectional Results by Prize or Duration

Variables	All	High Prize	Low Prize	Long Duration	Short Duration
	(1995 Cases)	Prize>287	Prize<287	Duration>21	Duration<21
N_Win	0.0108***	0.0268***	0.0034	0.0153***	0.0079**
G_SubmitOrder	- 1.3456**	- 0.4769	- 1.7922***	- 2.1892***	- 0.7137
G_SubmitOrder²	1.9449***	1.5790**	2.1074***	2.7392***	1.3767**
G_ΔTime	2.6157***	2.3481***	2.7777***	2.7633***	2.5140***
Pseudo R^2	8.58%	11.77%	7.08%	8.29%	8.98%

Note: *~p<0.1, *~p<0.05, ***~p<0.01.

Table 10. Correlations between Contest Designs and Expertise Signals

Expertise Signals		Ideation Based Projects				Expertise Based Projects				Ideation & Expertise Based Projects			
		Average	Bot-tier	Mid-tier	Top-tier	Average	Bot-tier	Mid-tier	Top-tier	Average	Bot-tier	Mid-tier	Top-tier
N_Join	A	-0.045	-0.019	-0.035	-0.035	-0.091	-0.057	-0.062	-0.136	-0.026	-0.032	-0.037	0.040
	D	**-0.167*****	-0.075	**-0.093***	-0.013	0.108	0.066	0.064	0.067	**-0.054****	-0.029	**-0.050***	**0.044***
N_Win	A	-0.076	-0.017	-0.021	-0.085	-0.099	-0.042	-0.058	**-0.149***	**-0.077*****	-0.032	**-0.044***	0.006
	D	**-0.200*****	-0.071	-0.087	**-0.173****	0.087	0.058	0.059	0.070	**-0.080*****	-0.032	-0.041	0.040
Award Won	A	**-0.119****	-0.017	-0.022	**-0.095***	-0.021	-0.001	-0.035	-0.031	-0.039	-0.021	-0.027	**0.053****
	D	**-0.178*****	-0.071	-0.081	-0.081	-0.058	0.036	0.059	-0.052	**-0.052****	-0.010	-0.031	**0.058****
Membership Age	A	0.082	-0.026	-0.001	0.147	-0.072	-0.081	-0.061	-0.044	0.013	-0.055	0.041	0.025
	D	**-0.110****	-0.097	**-0.097***	0.132	-0.145	0.028	0.036	-0.006	-0.040	-0.058	-0.030	-0.031
WinPerJoin	A	**-0.122****	-0.017	**-0.096***	-0.012	-0.011	0.011	-0.027	-0.037	**-0.110*****	**-0.061****	**-0.101*****	**0.069*****
	D	**-0.246*****	-0.071	**-0.180*****	0.079	**-0.230*****	-0.092	**-0.165***	**-0.170****	**-0.187*****	**-0.081*****	**-0.142*****	0.033
AwardWon/AwardJoined	A	**-0.147*****	-0.016	**-0.077***	0.017	0.003	0.013	-0.003	-0.019	**-0.092*****	**-0.054****	**-0.077*****	**0.053****
	D	**-0.247*****	-0.071	**-0.150*****	0.093	**-0.205****	-0.066	**-0.140***	**-0.141***	**-0.171*****	**-0.076*****	**-0.119*****	0.024

Note: *~p<0.1, *~p<0.05, ***~p<0.01, A~Award D~ Duration

To test H3, whether different contest designs can capture different groups of solvers, we calculate the correlations between contest designs (prize and duration) and all available expertise signals. In general, we calculate the correlations between average expertise of projects and contest designs. Besides, we also check the correlations according to different tiers of solvers. Since project dimensions may impact solver distribution, we put correlation results in three sections: ideation based projects, expertise based project, and ideation & expertise based projects according to dimensions defined in Table 5.

There are two additional signals we use in table 10: PrizeWon, and PrizeWon/PrizeJoined. PrizeWon is how much prize a solver has won in history, which is an experience measure and also is the "official" expertise indicator used by TaskCN. PrizeWon/PrizeJoined is another winning propensity measure in format of prize. PrizeJoined are the cumulative prize of all projects that a solver has participated. In total we have four experience based signals: N_Join, N_win, PrizeWon, and Membership Age. Besides, we have two winning propensity signals: WinPerJoin and (PrizeWon/PrizeJoined).

To interpret the correlation in table 10, for example, for ideation based projects, the correlation between solvers' average N_Join and duration is negative (ρ = - 0.167) and significant (p<0.01). That means for longer duration contest, the average N_Join of captured solvers is significantly lower. Similar result also holds for middle-tier solvers because the correlation is also negative and significant. However for top-tier solvers, or bottom-tier solvers, duration variation doesn't show significant association with N_Join.

In general, result shows that for ideation based projects (e.g. naming projects), the distributions of experience based signals and winning propensity signals are associated with prize or duration variations. And usually higher prize or longer duration results solvers with less experiences and lower winning propensity. Similar influence also holds for middle-tier solvers and top tier solvers. Although this is not expected by innovation seekers, since ideation based innovation is a highly random process and doesn't require specific expertise (Terwiesch and Xu 2008), it won't lower the individual performance significantly.

For expertise based projects (e.g. software development), it's surprising that prize has nearly no significant correlation with any expertise variation on average or any tier. For duration variation, only middle-tier and top-tier solvers' winning propensities are negatively and significantly associated. So for expertise based projects, assumption of constant expertise distribution is still hold if duration is fixed.

For ideation & expertise based projects (e.g. graphic design), the situation is similar to what we get for ideation based projects, where both experience based and winning propensity expertise are negatively and significantly correlated with the variation of prize or duration. In addition, winning propensity of bottom-tier solvers is also influenced. However, different from other two sections, correlations between prize/duration and expertise of top-tiers are positive and significant. That means, for higher prize or longer duration contests in this section, although captured solvers are with lower level of expertise on average, the top-tier solvers are with higher level of expertise. If the performance is mainly decided by the top-tier distribution (e.g. extreme value model), it's still beneficial for seekers to increase prize or duration.

Overall, our finding rejects H3 and suggests that the assumption of fixed expertise distribution is not always valid for online contest market.

Discussion and Conclusion

Our study provides empirical evidence that in an online contest market, both a solver's expertise and strategy are associated with his probability of winning. The number of winnings in past experience is a good predictor for future winning probability. Specifically, a solver with more winning experiences in the past will have higher probability of winning in future. In an open environment, earlier submitting solvers have the benefit of receiving feedback earlier from seekers while later submitters have the benefit of learning more information from competing submissions. Our results confirm that solvers with early submissions and late submissions are more likely to be winners than in-between submissions. We also show that those solvers who purposely "wait" in submission have higher probability of winning.

We also find that project characteristics can moderate expertise and strategy impact on probability of winning. For naming projects, higher prize and shorter duration will make the U shape probability curve flatter. And when prize is lower or duration is longer, strategic waiting doesn't help much in increasing probability of winning.

Finally, we find that solvers distribution is not constant across different projects. Solvers with more participation and winning experiences are not necessarily more likely to choose higher-prize projects. In fact, for ideation related projects, they show a trend of choosing relatively lower prize projects. Increasing prize is a common strategy that seekers are using, in order to attract better solutions. In general, our result shows they may actually reach lower quality solvers on average by doing that. However, for the most popular projects, which require both ideation and expertise, higher prize and longer duration can still help to capture better top-tier solvers. If the overall performance is decided by top-tier solver distribution, increasing prize or extending duration is still beneficial.

In general, why longer duration results lower quality solvers is unknown. Whether extending duration is beneficial is ambiguous. It seems an optimal duration exists. How to explain this phenomenon remains for future study.

In this study, while we have identified the importance of relative submission orders for solvers, it is still not clear how submission order is related to general submission time. From innovation seekers point of view, knowing when they will receive most and the best solutions is what they are more concerned about. It will be an interesting future study to explore factors that will expedite arrivals of good solutions. Moreover, our results have provided evidence that different projects, even within same category, attract solvers with different expertise level. Therefore, how to design a project that can attract better quality solvers, as of strategic concern to seekers, will be an interesting future research question too.

At last, the validity of our findings is depending on specific contest scenario. Once the fixed-prize policy or the visibility of submissions is changed, our results need to be revisited. It could be more interesting to use game theory and the bidding strategy and mechanism of the auction theory to explain the problems presented in this study.

References

Allio, R. 2004. "Ceo Interview: The Innocentive Model of Open Innovation," *Strategy & Leadership* (32:4), pp. 4-9.

Archak, N., and Sundararajan, A. 2009. "Optimal Design of Crowdsourcing Contests," *ICIS 2009 Proceedings*), p. 200.

Banker, R.D., and Iny, H. 2008. "Importance of Measures of Past experience: Empirical Evidence on Quality of E-Service Providers," *Contemporary Accounting Research* (25:2), Summer2008, pp. 307-337.

Brown, J.S., and Duguid, P. 1991. "Organizational Learning and Communities of Practice: Toward a Unified View of Working, Learning, and Innovation," *Organization Science* (2:1), pp. 40-57.

Faraj, S., and Sproull, L. 2000. "Coordinating Expertise in Software Development Teams," *Management Science* (46:12), p. 1554.

Howe, J. 2006. "The Rise of Crowdsourcing," *Wired Magazine* (14:6), pp. 1-4.

Liu, D., Geng, X., and Whinston, A. 2007. "Optimal Design of Consumer Contests," *Journal of Marketing* (71:4), pp. 140-155.

Ockenfels, A., and Roth, A. 2001. "Strategic Late-Bidding in Continuous-Time Second-Price Internet Auctions,").

Ockenfels, A., and Roth, A.E. 2006. "Late and Multiple Bidding in Second Price Internet Auctions: Theory and Evidence Concerning Different Rules for Ending an Auction," *Games and Economic Behavior* (55:2), pp. 297-320.

Rorty, R. 1979. "Philosophy and the Minor of Nature," *Princeton UP*).

Roth, A., and Ockenfels, A. 2002. "Last-Minute Bidding and the Rules for Ending Second-Price Auctions: Evidence from Ebay and Amazon Auctions on the Internet," *American Economic Review* (92:4), pp. 1093-1103.

Snir, E., and Hitt, L. 2003. "Costly Bidding in Online Markets for It Services," *Management Science* (49:11), pp. 1504-1520.

Terwiesch, C., and Ulrich, K. 2009. *Innovation Tournaments: Creating and Selecting Exceptional Opportunities*. Harvard Business School Pr.

Terwiesch, C., and Xu, Y. 2008. "Innovation Contests, Open Innovation, and Multiagent Problem Solving," *Management Science* (54:9), pp. 1529-1543.

Vadovic, R. 2009. "Early, Late, and Multiple Bidding in Internet Auctions," *Working Papers*).

Walsh, J.P. 1995. "Managerial and Organizational Cognition: Notes from a Trip Down Memory Lane," *Organization Science* (6:3), pp. 280-321.

Yang, J., Adamic, L., and Ackerman, M. 2008. "Crowdsourcing and Knowledge Sharing: Strategic User Behavior on Taskcn," ACM, pp. 246-255.

Yang, Y., Chen, P., and Pavlou, P. 2009. "Open Innovation: An Empirical Study of Online Contests," *ICIS 2009 Proceedings | International Conference on Information Systems*), p. 13.